

Suggested Method to Adjust the Parameters of SVM Regression using Particle Swarm Optimization

Aseel Sameer Mohamed

¹ University of Baghdad, Al Kindy College of Medicine, Community & Family Medicine Department

aseelsameer@kmc.uobaghdad.edu

ARTICLE INFO

Received: 18 Dec 2024

Revised: 10 Feb 2025

Accepted: 10 Apr 2025

ABSTRACT

The Gaussian Process Regression (GPR) was first introduced as a regression tool in the field of machine learning by Rasmussen and Williams in 1996, where they described the optimization of parameters in the covariance function, which was inspired by the use of the Gaussian process with neural networks.

This research aims primarily to study and improve the performance of the Gaussian process regression by reaching the best method for adjusting the values of the coefficients of this technique.

In this research, a new method was proposed for testing the adjustment of the values of the Gaussian process regression coefficients grounded on the Particle Swarm Optimization (PSO) algorithm, with the aim of reaching the best method for choosing the values of these coefficients.

In order to apply the two methods, a script was written in a statistical programming language (R), which performs many data analysis by using Packages.

Addicts is an important social aspects and dangerous phenomenon in population because its aim is to destroy the minds and bodies of young people and at the same time destroy them, therefore the phenomenon of drug addiction has become one of the serious problems that preoccupy officials in all parts of the world, especially our Islamic world, and day after day the danger of addiction is exacerbated, because it increases every day with the decrease in the age of addiction.

Realistic data was collected from the Ibn-Rushed Psychiatric Hospital in Baghdad , involved 196 males from July 2023 to February 2024 involved 196 males, with addiction durations with varying degrees of methamphetamine (METH) addiction, which represent the duration and four test measure which is commonly used to help diagnose liver damage or disease which are Alanine aminotransferase test (ALT) , Aspartate aminotransferase test (AST) ,Alkaline Phosphatase (ALP) , and Gamma – glut amyl Transfers (GGT).

The following measures (MSE, RMSE, MAPE) are conceded the best measures for comparing different prediction models built using the same training dataset (Hyndman and Koehler, 2006)..

Keywords: Addicts Males, Gaussian Process Regression), Particle Swarm Optimization, Cross correction.

INTRODUCTION

The Gaussian Process Regression (GPR) was first introduced equally a regression tool in the machine learning area of by Rasmussen and Williams in 1996, where they described the optimization of parameters in the covariance function, which was inspired by the use of the Gaussian process with neural networks. It has been used in various applications such as prediction of skin permeability of chemicals and predicting the concentration of ozone in the air (Rasmussen 2007, Bishop and Williams 2006).

Most algorithms or methods for selecting GRP parameters depend on experience and random searches, and despite the development of search algorithms, there are still some disadvantages to applying these algorithms to this issue, as they may limit the quality of performance and accuracy of prediction, so this issue needs further study (Bishop 2007).

The importance of the research lies in developing prediction methods as they create the basic building block of the treating data and knowledge extraction process. We cannot build knowledge on the basis of predictions that do not represent the provided databases well, and the applications of these methods are many it is involved in various fields of science and its development helps in the progress of scientific research.

This research aims primarily to study and improve the performance of the Gaussian process regression by reaching the best method for adjusting the values of the coefficients of this technique.

The problem of choosing the basic Gaussian process regression coefficients lies in the absence of a general methodology for adjusting the values of these coefficients, due to the difficulty of studying the issue analytically.

2. MATERIALS AND METHODS

2.1. Gaussian Process (GP)

The Gaussian process regression model is a nonparametric model, i.e. it does not accept a specific form of the purpose under study, but the procedure of the association amid the inputs and the objectives is completely strong-minded by the data, which may contain an limitless amount of functions. The basic purpose that crops the data is indefinite, but the predictions are generated by a set of functions that are topic to a Gaussian distribution in the function space. It is considered the Gaussian process regression model is one of the latest prediction methods, and it is one of the Bayesian probabilistic methods, and it is one of the Bayesian probabilistic models. In most Bayesian regression methods Prior information about the model parameters is found, and then conditions are placed on the data to give the model posterior parameters. This prior information can be formulated in the form of a probability distribution called the prior distribution. The Bayesian model specifies the unknown parameters of the prior model, while the Gaussian process model specifies the relationships of the functions. Direct correlation between test inputs and training inputs and outputs (Rasmussen and Liu 2006, Williams et al., 2017).

A group of random variables with a joint Gaussian distribution that is limited in number is called a Gaussian process. In probabilistic machine learning, Gaussian Processes have emerged as a powerful tool that provides a flexible and ethical framework for forming uncertainty in a range of applications. (Rasmusse & I. Williams, 2006, 40)

Let X be a set, $(X, X') : X \times X \in \mathbb{R}_+$ be a kernel and $\mu_X : X \in \mathbb{R}_+$ real-valued function. Then a random function $f : X \in \mathbb{R}$ is said to be Gaussian process (GP) with mean function μ_X and Kernel function $K(X, X')$, denoted by $GP(\mu_X, K(X, X'))$ if any finite set $X = (X_1, X_2, \dots, X_n)$ of size $n \in \mathbb{N}$. (Shi et al., 2022, 41733) (Kanagawa et al. 2018, 8). Then,

$$f(X_{ij}) = (f(X_1), f(X_2), \dots, f(X_n)) \in \mathbb{R}^n \dots (4)$$

have the multivariate normal distribution $MN(\mu_X, K(X, X'))$, where: $K(X, X') = (k(x_i, x_j))_{i,j=1}^n \in \mathbb{R}^{n \times n}$ is Kernel function and $\mu_X = (\mu(x_1), \mu(x_2), \dots, \mu(x_n))$ is mean vector. and,

$$K(X, X') = E[f(X_{ij}) - \mu_X] [f(X_{ij}) - \mu_X]^T \dots (5)$$

Where $\mu_X = E[f(X_{ij})]$ where $i = 1, 2, \dots, n, j = 1, 2, \dots, q$

Then the probability distribution for GP for (n) observation and q explanatory variables X_{ij} as following,

(J. L. Akian et al., 2022)

$$\begin{pmatrix} f(X_1) \\ f(X_2) \\ \vdots \\ f(X_n) \end{pmatrix} \sim \text{MVN} \left[\begin{pmatrix} \mu(X_1) \\ \mu(X_2) \\ \vdots \\ \mu(X_n) \end{pmatrix}, \begin{pmatrix} K(X_1, X_1') & K(X_1, X_2') & \dots & K(X_1, X_n') \\ K(X_2, X_1') & K(X_2, X_2') & \dots & K(X_2, X_n') \\ \vdots & \vdots & \ddots & \vdots \\ K(X_n, X_1') & K(X_n, X_2') & \dots & K(X_n, X_n') \end{pmatrix} \right] \dots (6)$$

(H. Liu J. Cai, 2020, 12) (اكو تكملة مال كيرنل)

$$P(f(x)) = GP(\mu(x), k(x, x')) \quad \text{--- (7)}$$

Where $k(x, x')$ is called the covariance function or the kernel function which is a finite positive function and has several types. (Bishop, 2007).

Let the nonparametric regression model, as in equation (8)

$$y \sim GP(o, k(x, x')) + \sigma_n^2 \delta(x, x') \quad \text{--- (8)}$$

$$Y = f(x) + \epsilon \quad \epsilon \sim (0, \sigma^2)$$

Where Y be a dependent variable and X be random variables with d dimensions, $f(x)$ in a known function in parametric regression .

Let a set of data $\{(x_i, y_i)\}_{i=1}^n$ where x_i represent the inputs which have a dimension d , while y_i represent the true values for products , n represent the number of data.

The model of Gaussian process regression is as in equation (9)

$$Y_i = f(x_i) + \epsilon_i \quad i = 1, 2, \dots, n \quad \epsilon \sim (0, \sigma^2) \quad \text{----- (9)}$$

Where $f(x) \sim GP(\mu(x), k(x, x'))$, $GP(\mu(x), k(x, x'))$ is prior Gaussian with $\mu(x)$, and kernel function $k(x, x')$, so the model will be as in equation (10)

$$y = GP(\mu(x), k(x, x')) + \sigma_n^2 \delta(x, x') \quad \text{----- (10)}$$

Where $\delta(x, x')$ is Kronecker delta , $\delta(x, x')=0$ when $x \neq x'$, $\delta(x, x')=1$ when $x = x'$, σ_n^2 is random ضجيج

$$k(x, x') = e^{-(||x - x' ||)/2\sigma^2) \quad \text{----- (11)}$$

Where $||x - x' || = \sqrt{(x - x')^T (x - x')}$ denotes the long ray or the normalized difference.

(Rasmussen and Williams, 2006).

2.2. Analytical method for adjusted the Gaussian process regression Coefficients:

The analytical or experimental selection method is based on adjusting the value of random noise coefficient σ_n and the kernel function coefficient σ directly from the training data .This method is characterized by its ease of calculation and gives acceptable results. The value of the coefficient and the kernel function coefficient are given , according to the equation (12)

$$\sigma = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X})^2} \quad \text{---- (12)}$$

So the expected values of y are the regression values by the nearest neighbor method and the coefficient k represent the number of neighborhood points in the nearest neighbor algorithm which is always be $\{3,5,7\}$. (Bishop , 2007 ; Rasmussen & Williams 2006; Sh. & Choi 2011).

In this research we propose and set the value of k to 5 to use the long set Euclidean distance between the values of x to represent the data range that cannot be calculated in the multidimensional case.

2.3. The Proposed method for adjust the Gaussian Process regression coefficients :

The selection of hyper- parameters in the divergent process regression technique plays an important role in obtaining a good model (Shi & Choi 2011).

In this research , we proposed using the Particle Swarm Optimization algorithm in choosing and adjusting the values of the Gaussian process regression coefficients in order to reach the best method for adjusting the values of these coefficients.

Particle Swarm Optimization (PSO) recorded advanced algorithms problem in common . It mimics the normal behavior initiate in flocks of birds of schools of fish while searching for food sources (Aich et al .2014, Banerjee et al . 2010).

In this algorithm , a random initial set of elements called a swarm is taken and the elements of this swarm update their positions and velocities based on the information collected by swarm .

Each element, in addition to the effect of the entire swarm, influences the direction of the entire swarm towards the optimal region.

The convergence rate is controlled using several parameters , which are :

- 1- Inertia Factor(w) controls the effect of the previous velocity vector of an element on the current velocity vector of this element .
- 2- Construction Factor determines the effect of the velocity vectors of neighboring elements on the current position of the element and aims to avoid the clustering of elements around one point in the search space to avoid falling into local optimal solutions.
- 3- Cognitive Acceleration c_1 , which controls the effect of the experience of individuals through the best location that this element has passed through.
- 4- Social Acceleration factor c_2 , which controls the effect of the entire swarm through the best location that an element has reached comprehensively so far (Aich and Banerjee , 2014) . Choosing the number of swarm elements is also very important . When choosing a small number of swarm elements , we will not be able to collect sufficient information about the entire search area, and choosing a large number of swarm elements will lead to consuming more time to converge from the optimal solution area. It should be noted that the importance of the random distribution of the initial swarm elements of spread over the entire area the assumed search is to cover the entire search area, and the maximum number of iterations is chosen based on the experiment , it is important to reach convergence otherwise we will not get good results , and the iterations that follow reaching the solution are unless because the equilibrium point is reached and the velocity rays for all elements are zero and the maximum number of iterations is set as a condition for stopping (Mullen 2014).

2.3.1. Steps of the PSO algorithm to choose the GPR (σ, σ_n) :

- First : Determine the areas of internal control coefficients and the standard values were used within the software package used and the number of swarm elements was determined and set in this search to $m=30$ and the maximum number of iteration $iter_{max} = 100$.
- Second : For the first $iter=1$, the initial position vector values of the swarm elements are generated randomly within the search area ranges so that each ray represents the position of the GPR coefficients (σ, σ_n) and the initial velocity vector for each element is calculated in equation (13)

$$V_{iter}^{(j)} = rang^{(j)} * rand \quad j = 1,2,3 \quad (13)$$

Where $rang^{(j)}$ refers to the coefficient ranges on (σ, σ_n) and $rand$ refers to a random number within the range $[0,1]$ and j indicates that the calculation will be repeated three times to generate the initial velocity for the GPR (σ, σ_n) and thus the velocity vector m is generated with the components corresponding to the generated with the components corresponding (σ, σ_n) to be calculated.

- Third: Set $t=1$ and $Pbest\ t$, $Gbest$ equal to the current position vector .
- Fourth: If $t=m$, we move to step number 8 , otherwise we set $t=t+1$ and move to step 5.
- Fifth: Calculate the three error measures (MAPE, RMSE, MSE) used at position t in the swarm and in case $t=1$, we assign $Pbest\ t$ to the current position vector and move to step 7 , otherwise we move to step number 6.
- Sixth: If the value of the error scale used is better than the error scale of the current position $Pbest\ t$ ray with the current position ray , otherwise the $Pbest\ t$ values remains the same .
- Seventh: If the value of the error scale used is better than the error scale of the current position $Gbest$, we update the $Gbest$ ray with the current position ray , otherwise the $Gbest$ value remains the same , and in both cases we move to step 4.
- Eighth: We reset the internal control parameters for the current iteration number $iter$ by the following equations :

$$W_{iter} = W_{initial} + (W_{final} - W_{initial}) (iter - 1) / (iter_{max} - 1)$$

$$Y_{iter} = Y_{initial} + (Y_{final} - Y_{initial}) (iter - 1) / (iter_{max} - 1)$$

$$C_{1,iter} = C_{1,initial} + (C_{1,final} - C_{1,initial}) (iter - 1) / (iter_{max} - 1)$$

$$C_{2,iter} = C_{2,initial} + (C_{2,final} - C_{2,initial}) (iter - 1) / (iter_{max} - 1)$$

So that itermax is the maximum number of iterations , and the velocity vector for all elements are updated using the equation (14):

$$V_{(j),iter+1}^k = w_{iter} (V_{(j),iter}^k) + c_1 (rand) (Pbest_{(j)}^k - P_{(j),iter}^k) + c_2 (rand) (Pbest_{(j)}^k - P_{(j),iter}^k) \quad j=1,2,3 \quad \text{---(14)}$$

So that $V_{(j),iter}^k$ is the velocity vector of the component of the coefficients j from the element k in iteration number $iter$ and so that $P_{(j),iter}^k$ is the position vector of j for the element k in iteration $iter$ number $iter$ and so that j counter repeats the calculations three times to generate the velocity vector for the GPR coefficients.

It is worth noting that the previous relations can be divided into a total of three terms so that the first term represents the Memorial Component and the second term represents the cognitive component while the third term represent the social component , meaning that the new velocity vector is affected by the three factors memory , acquired knowledge , and swarm intelligence (Hu,2010 , and his colleagues).

The position vectors are updated for all swarm elements using the equation (15)

$$j=1,2,2 \quad \text{---(15)} \quad P_{(j),iter+1}^k = P_{(j),iter}^k + V_{(j),iter+1}^k$$

In case the new position of the components corresponding to the coefficients σ , σ_n) is outside the specified search area, the current position is returned to the borders of the search area.

- Ninth: If $iter < itermax$, we put $iter = iter + 1$ and put $t=1$ and we move to fifth step. In the case of $iter = itermax$, the current value of Gbest is adopted as optimal solution , meaning that the components of the Gbest ray have become the optimal values for the n coefficients, which makes the value of the error scale used as small as possible when applying the GPR algorithm to the training data .

3. APPLIED ASPECT:

To exam the precise of the proposed method in adjusting the Gaussian process regression (σ , σ_n) , we applied the AS analytical method and the PSO element swarm algorithm method to Iraqi males suffering from addiction.

Addicts is an important social aspects and dangerous phenomenon in population because its aim is to destroy the minds and bodies of young people and at the same time destroy them , therefore the phenomenon of drug addiction has become one of the serious problems that preoccupy officials in all parts of the world , especially our Islamic world , and day after day the danger of addiction is exacerbated , because it increases every day with the decrease in the age of addiction .

Realistic data was collected from the Ibn-Rushed Psychiatric Hospital in Baghdad , composed 196 males from July 2023 to February 2024, with varying grades of methamphetamine (METH) addiction, which represent the duration intake and four test measure which is commonly used to help diagnose liver damage or disease which are Alanine aminotransferase test(ALT) , Aspartate aminotransferase test (AST) ,Alkaline Phosphatase (ALP) , and Gamma – glut amyl Trasferase(GGT).

The prediction model to be built aims to estimate the degrees of methamphetamine (METH) by knowing the liver function tests (ALT, AST, ALP, and GGT) , so that the duration intake represent the dependent variable while the liver function represent the impendent variables respectively x_1 to x_4 .

3.1. Proposed Methodology Testing Methodology:

Accuracy is a criterion for choosing the optimal prediction model. Accuracy refers to the ability of the prediction model to reproduce the original data of the studied sample. Therefore, the appropriate choice of the prediction accuracy measure has a positive effect on determining the effectiveness of the used prediction model.

Standard prediction accuracy measures generally work on the concept of the difference between the original data and the expected or predicted data, which is the predicted error. The smaller the difference , the better and more accurate the prediction function .There are many prediction accuracy measures, and usually one measure is not relied upon in the prediction model tuning process.

The following measures (MSE, RMSE, MAPE) are conceded the best measures for comparing different prediction models built using the same training dataset (Hyndman and Koehler, 2006).

In this research, we relied on these measures because they suite the nature of the research and they calculated as follows:

In this research , we relied on these measures :

Mean Squared Error MSE= mean (e_i^2)

Root Mean Squared error RMSE= \sqrt{MSE}

Mean Absolute percentage error MAPE = mean ($|e_i/y_i|$)

Where residuals $e_i = y_i - \hat{y}_i$ (Hardle et al ,2004).

To evaluate the quality of performance of each prediction method, the values of these measures were calculated in two ways: the first consuming the entire training data set , and the second by applying the fold Cross Validate methodology to measure the precision of the resulting model using the coefficients calculated in each method so that this methodology divides the data into a section .One of the sections is chosen each time as a test sample and the rest as training samples. The error measure is calculated using the test sample time .Finally, the average of the error measure values obtained in each cycle are calculated . The value is usually set to one of the values 3 , 5 or 10 (Hyndman Koehler , 2006), and in this research select k=10.

In order to apply the two methods, a script was written in the R programming language which is perform many data analyses so that theses analyses are organized within Packages , which means that researchers are able to develop different programs , which contributed to the spread of their use in academic fields (Cotton , 2033; Matloff , 2011).

To implement this research , the following package is used :

The caret package was used to apply the nearset neighbour algorithm to estimate the noise value and calculate the Euclidean distance matrix to calculate the Gaussian process regression coefficients using the analytical method , the kernlab package was used to apply the Gaussian process regression , and the hydroPSO package was used to apply the element swarm optimization algorithm. The cross-correction methodology was programmed, and no package was used to apply more than one error measure.

4. RESULTS & DISCUSSION:

After the practical application, as in Table (1):

Table (1) : Values of the coefficients

Methods	Coefficients Values	
	σ_n	σ
Analytical Method	0.2036204	108.7684
Proposed Method	0.04710621	0.20919407

After applying the analytical method and the proposed method according to the coefficients shown in Table (2) on the addict's data set and calculating the error scale values as in Table (2):

Measures	Mean Squared Error		Mean Absolute percentage error		Root Mean Squared Error	
Methods	Classical	Proposed	Classical	Proposed	Classical	Proposed
10-cv	0.6212961	0.7441916	0.3065102	0.1621071	0.7882233	0.8626654
Cross Correction						
All Data	0.4643556	0.6761001	0.2502411	0.08464622	0.6814364	0.8222530

From Table (2) we note that the proposed method had smaller values for the three measures (MSE , RMSE, MAPE) , so we notice the effectiveness of the proposed PSO algorithm in estimating GPR coefficients due to its ability to converge and reach the optimal solution with the least number of iterations, as it converged from the best fit in step 13, so we relied to measure the speed of algorithm convergence to the solution on the algorithm execution time in seconds in addition to the number of convergence iterations is set to 50 iteration in which the analytical method does not need any iteration .

5. CONCLUSION AND RECOMMENDATIONS:

- 1- The study showed the super priority of the proposed method represented by using the element swarm optimization algorithm PSO in adjusting the Gaussian process regression coefficients GPR over the experimental analytical method by achieving the smallest value of the three error measures (MSE, MAPE, RMSE) whether using the entire set of data to calculate the errors or using the cross- correction methodology.
- 2- Through the results of this study, the effectiveness of the proposed element swarm optimization algorithm PSO in adjusting the GPR coefficients due to its ability to converge and reach the optimal solution.
- 3- It was also found that PSO algorithm gives the best prediction accuracy , the best result , and the fastest convergence to the solution due to the repetition of the priority number built on the minimum value of the error measures used .
- 4- We recommended using the proposed PSO algorithm in adjusting the GPR coefficients as it gives the best results.
- 5- We recommended programming this method using parallel programming so that many operations are performed concurrently , which is founded on the value of big complications can frequently be separated into minor difficulties to be solved in parallel at the same time to reduce the time required to implement them.
- 6- We recommend adopting additional stopping conditions when there is no improvement in the fit value after specific number of iterations to benefit from the convergence speed property in the PSO algorithm.
- 7- Conduct future studies on adjusting the GPR parameters using methods other than the method proposed in this research.

REFERENCES:

- [1] Aich, U., and Banerjee, S., (2014) “ Modeling of EDM Responses by Support Vector Machine Regression with parameters Selected by Particle Swarm Optimization “ Applied Mathematical Modeling , 38, pp.: 2800-2818.
- [2] Bishop, C. M. , (2007) “ Pattern Recognition and Machine Learning “ Springer.
- [3] C. Benedek, A. Majdik, B. sNagy, Z. Rozsa, T. Sziranyi(2021) “ Positioning and perception in LIDAR point clouds” , Digit. Signal Process. 119.
- [4] Cotton, R., (2013) “ Learning R” O Reilly Media , Inc., United States of America, 377.
- [5] F. Bu, T. Le, X. Du, R. Vasudevan, R. Vasudevan, M. Johnson-Roberson,(2020)” Pedestrian planar LiDAR pose (PPLP) network for oriented pedestrian detection based on planar LiDAR and monocular images” , IEEE Robot. Autom. Lett. 5 pp: 1626–1633.
- [6] Hardle, W., Muller, M. , Sperlich , S., and Werwatz A., (2004) “ Nonparametric and Semi parametric Models “ Springer , Berlin , pp:301.
- [7] Hu, Z.,Min, W , Huang , X., (2010) “ Parameter Selection of Support Vector Regression Basred on Particle Swarm Optimization” International Conference on Granular Computing , 5 pp:251-256.
- [8] Hyndman, R.,J.,Koehler , A.,B., (2006) “ Another lock at Measures of Forecast Accuracy “ International Journal of Forecasting , 22 pp:679-688.
- [9] Liu,Y.,Keller, Y., Song, PH., Bond , J., Jiang , G. (2017) “ Prediction of concrete corrosion in sewers with hybrid Gaussian processes regression model “ RSC Advances , 7, pp:30894-30903.
- [10] Matloff, N., (2011) “ The Art of R Programming , Malloy Incorporated “ United States of America, pp:373.
- [11] M Hasan, J Hanawa, R Goto, R Suzuki, H Fukuda, Y Kuno, Y Kobayashi(2022) ” LiDAR-based detection, tracking, and property estimation: A contemporary review” Neurocomputing , Elsevier.
- [12] Mullen , K., (2014) “ Continous Global Optimization in R” Journal of Statistical Software , 60 Issue 6.
- [13] Rasmussen, C.,E., Williams C.K. I., (2006)” Gaussian Processes for Machine Learning “ MIT , Press.
- [14] R. Wang, M. An, S. Shao, M. Yu, S. Wang, X. Xu,(2021)” Lidar sensor-based object recognition using machine learning” , J. Russ. Laser Res. 42 PP: 484–493.
- [15] Shi,J.,Q., Choi, T., (2011)” Gaussian Process Regression Analysis for Functional Data “ Chapman& Hall/CRC” London , rst edition .
- [16] Y. Wu, Y. Wang, S. Zhang, H. Ogai,(2021) “ Deep 3D object detection networks using LiDAR data: A review” , IEEE Sens. J. 21, pp: 1152–1171.
- [17] Y. Yao, S. Fatholahi, Y. Chen, J. Li(2023)” Indoor LiDAR point clouds upsampling for object detection enhancement” , in: Joint Urban Remote Sensing Event, JURSE.